Klaus Pötzelberger

Admissible Unbiased Quantizations: Distributions without Linear Components

Working Paper

Admissible Unbiased Quantizations: Distributions without Linear Components

Klaus Pötzelberger

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 76
June 2000

http://statmath.wu-wien.ac.at/

Admissible Unbiased Quantizations: Distributions Without Linear Component

Klaus Pötzelberger

Abstract

Let P be a Borel probability measure on $\mathbb{R}^d$. We characterize the maximal elements $\mu \in \mathcal{M}(P,m)$ with respect to the Bishop-De Leeuw order $\preceq$, where $\mu \in \mathcal{M}(P,m)$ if and only if $\mu \preceq P$ and $|\text{supp}(\mu)| \leq m$. The results obtained have important consequences for statistical inference, such as tests of homogeneity or multivariate cluster analysis and for the theory of comparison of experiments.

AMS 1991 subject classifications. Primary 62C15; secondary 62B10; 62B15, 62H30.
Key words and phrases. Admissibility, quantization, Bishop-de Leeuw order, dilation, information, majorization, partitions, MSP-partitions.

1 Introduction

Let $E = (\Omega, \mathcal{F}, (P_1, \ldots, P_d))$ denote an experiment of order d. The complexity of the experiment may be reduced by replacing $\mathcal{F}$ by a finite field $\tilde{\mathcal{F}} \subseteq \mathcal{F}$, which leads to the experiment $F = (\Omega, \tilde{\mathcal{F}}, (P_1, \ldots, P_d))$. $\tilde{\mathcal{F}}$ is identified with a finite partition $\mathcal{B} = (B_1, \ldots, B_m)$ of Ω. Recall the concept of information for experiments. The results obtained in this paper give a characterization of the reduced experiments which are maximal with respect to the information semiorder on the set of all reduced experiments with the size of the corresponding partition being at most m.

Let us briefly provide examples from statistical applications. Consider a probability space $(\Omega, \mathcal{F}, P)$. Various statistical methods comprehend the approximation of P by a distribution μ with finite support. This is a task related to the choice of a finite algebra $\mathcal{F}_0 \subseteq \mathcal{F}$, i.e. a partition of the sample space, with minimal loss of information.

For instance, in descriptive statistics, quantities such as principal points in the sense of Flury [4] or quantiles are assigned to distributions. Continuous laws are replaced by discrete laws by rounding or grouping. In cluster analysis an empirical distribution, i.e. data, is partitioned such that some measure of homogeneity is maximized within and minimized between clusters.

Procedures based on a partition of the sample space are common in inference statistics. Think of the χ^2-test of homonegeity. Here typically the distribution of metrically scaled random variables is replaced by a multinomial distribution. The power of the test depends on the chosen partition of the sample space. See Bock [3] for details.

In all these procedures the grouping of data leads to a loss of information. In the majority of cases μ is chosen from a specified class of distributions in order to maximize a given measure of information. Let us illustrate this for principal points. Suppose P is a distribution on $\mathbb{R}^d$ with finite second moment. For a partition $\mathcal{B} = (B_1, \ldots, B_m)$ define the conditional means

$$p_i = \int_{B_i} x dP/P(B_i). \tag{1}$$

A partition $\mathcal{B}$ is optimal, if it maximizes the information measure

$$\sum_{i=1}^{m} \|p_i\|^2 P(B_i)$$

among all partitions of size at most m. The conditional means (p_i) are called principal points or prototypes. Let $f(x) = \|x\|^2$. Note that the discrete distribution

$$\mu = \sum_{i=1}^{m} P(B_i)\delta_{p_i}, \tag{2}$$

corresponding to $\mathcal{B}$, maximizes $\mu(f)$ if and only if $\mathcal{B}$ is optimal. δ_x denotes the Dirac distribution in x.

A further example is related to the algorithm of Kohonen [5]. Here a partition $\mathcal{B}$ is sought such that μ, defined by (2), maximizes $\mu(f)$, with convex function $f(x) = \|x\|$.

Pötzelberger and Strasser [7] analyze general procedures of this type. Let P be a Borel distribution on $\mathbb{R}^d$ with $\int \|x\| dP < \infty$ and $P(H) = 0$ for all hyperplanes H. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function with $P(f) < \infty$. Let $\mathcal{B} = (B_1, \ldots, B_m)$ be a partition and let the prototypes and p_i and μ be given by (1) and (2). Define

$$I^f(\mathcal{B}) = \mu(f) \tag{3}$$

and

$$I^f_m = \sup\{I^f(\mathcal{B}) \mid |\mathcal{B}| \leq m\}. \tag{4}$$

Let us call $I^f(\mathcal{B})$ the information of the partition $\mathcal{B}$ (or of the distribution μ). Consider L_m, the set of convex functions that are piecewise linear and linear on m sets. I.e. $g \in L_m$ if and only if $c_1, \ldots, c_m \in \mathbb{R}$ and $d_1, \ldots, d_m \in \mathbb{R}^d$ exist such that

$$g(x) = \max\{c_i + \langle x, d_i \rangle \mid 1 \leq i \leq m\} \qquad P - a.e. \tag{5}$$

$\langle .,. \rangle$ denotes the scalar product on $\mathbb{R}^d$. $g \in L_m$ generates the partition $\mathcal{B} = (B_1, \ldots, B_m)$ if $B_i \subseteq \{x \mid g_i(x) \geq g_j(x) \text{ for all } j\}$, i.e. g is linear on B_i. Partitions generated by $g \in L_m$ are called Maximal-Support-Plane partition (MSP-partition). MSP-partitions of the real line have been studied by Bock [3] in the context of quantizing likelihood ratios.

Let f be convex and $p \in \mathbb{R}^d$. Denote by $D(f,p)$ the set of subdifferentials of f in p, i.e. $d \in D(f,p)$ if $g \leq f$, where the linear function g is defined by $g(x) = f(p) + \langle x - p, d \rangle$. Let $p_1, \ldots, p_m \in \mathbb{R}^d$, $d_i \in D(f, p_i)$ and $g_{p_1,\ldots,p_m,d_1,\ldots,d_m}(x) = \max\{f(p_i) + \langle x - p_i, d_i \rangle \mid 1 \leq i \leq m\}$. Pötzelberger and Strasser [7] have shown that the choice of an optimal partition $\mathcal{B}$ is equivalent to the choice of prototypes $p_1, \ldots, p_m$ and subdifferentials $d_1, \ldots, d_m$ which maximize $P(g_{p_1,\ldots,p_m,d_1,\ldots,d_m})$. More specifically, let $\mathcal{B}$ be a partition with $|\mathcal{B}| \leq m$ and $I^f(\mathcal{B}) = I^f_m$. If the prototypes p_i are defined by (1) and $d_i \in D(f, p_i)$, then $g_{p_1,\ldots,p_m,d_1,\ldots,d_m}$ maximizes the functional $P(g)$ among all $g \in L_m$ with $g \leq f$. On the other hand, if $P(g)$ is maximized among all $g \in L_m$ with $g \leq f$, then the partition $\mathcal{B}$ generated by g is optimal, i.e. $I^f(\mathcal{B}) = I^f_m$.

Thus for given convex function f, an optimal partition is automatically a MSP-partiton. For the choice of the function f conceptual as well as methodological considerations, such as robustness of the procedure, are relevant (cf. Pötzelberger and Strasser [7]).

The concept of information of partitions or quantizations is based on the comparison of expectations of convex functions. This concept leads to an order on the set of probability distributions, the dilation or Bishop-De Leeuw order. Consider the set of distributions with size of the support fixed and which are dominated by P in the Bishop-De Leeuw order. It is the aim of this paper to provide a complete characterization of distributions which are maximal with respect to the Bishop-De Leeuw order in this set of distributions.

DEFINITION 1. *Let us define the Bishop-De Leeuw order $\preceq$ for probability distributions. Let P and Q be Borel distributions on $\mathbb{R}^d$. $Q \preceq P$ if a stochastic kernel $K(dx \mid y)$ exists such that*

$$\int x K(dx \mid y) = y \tag{6}$$

and $P = KQ$, i.e. for all Borel sets A

$$P(A) = \int K(A \mid y) Q(dy). \tag{7}$$

DEFINITION 2. *Let P be a Borel distribution on $\mathbb{R}^d$ and $m \in \mathbb{N}$. We define*

$$\mathcal{M}(P, m) = \{\mu \mid \mu \preceq P \text{ and } |supp(\mu)| \leq m\}. \tag{8}$$

Let us call the elements of $\mathcal{M}(P, m)$ unbiased quantizations of P. A maximal element of $\mathcal{M}(P, m)$ is called an admissible quantization, i.e. $\mu \in \mathcal{M}(P, m)$ is admissible if $\nu \in \mathcal{M}(P, m)$ and $\mu \preceq \nu$ imply $\mu = \nu$.

REMARK 1. (a) If $|\mathrm{supp}(P)| \geq m$ and $\mu \in \mathcal{M}(P, m)$ is admissible, then $|\mathrm{supp}(\mu)| = m$.

(b) The Theorem of Blackwell-Sherman-Stein (see Torgersen [12] or Strassen [10]) provides the fundamental characterization of the Bishop-De Leeuw order: $Q \preceq P$ if and only if for all convex and continuous functions f

$$Q(f) \leq P(f). \tag{9}$$

(c) Let $\mu = \sum_{i=1}^{m} w_i \delta_{p_i}$ be a distribution on $\mathbb{R}^d$ with $|\text{supp}(\mu)| \leq m$. μ is an unbiased quantization of P if Borel distributions P_i exist such that

$$\int x P_i(dx) = p_i \tag{10}$$

and

$$P = \sum_{i=1}^{m} w_i P_i. \tag{11}$$

Let us briefly mention the significance of the Bishop-De Leeuw order and the concept of majorization in the theory of comparison of experiments. This theory goes back to Blackwell ([1], [2]). For details see Strasser [11] or Torgersen [12]. Consider two finite experiments $E = (\Omega_1, \mathcal{F}_1, (P_1, \ldots, P_d))$ and $F = (\Omega_2, \mathcal{F}_2, (Q_1, \ldots, Q_d))$ of the same order d. Experiment E is more informative than F ($E \supseteq F$) if a stochastic kernel K from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ exists with $Q_i = KP_i$ for $i = 1, \ldots, d$. Suppose $\mathcal{F}_2$ is a finite subfield of $\mathcal{F}_1$ and $F = (\Omega_1, \mathcal{F}_2, (P_1, \ldots, P_d))$. Then $E \supseteq F$. The decision theoretic aspect of quantization is the choice of the finite subfield $\mathcal{F}_2$ generated by a partition of given size. Let us define for a finite experiment E the standard measure P_E. Let $\bar{P} = \sum_{i=1}^{d} P_i$ and let P_E be the law of the likelihood ratios $(dP_i/d\bar{P})_{i=1}^{d}$ under the law $\bar{P}/d$. A theorem of Blackwell [1] states that $E \supseteq F$ if and only of $P_F \preceq P_E$. Thus procedures based on quantizations of the standard measure which are inadmissible are dominated by procedures based on admissible quantizations.

2 Results

A set $H = \{x \mid \langle x, b \rangle = a\}$ is called a hyperplane. A halfspace is a set $E = \{x \mid \langle x, b \rangle \geq a\}$. a is a scalar and $b \in \mathbb{R}^d$.

DEFINITION 3. *Let P be a probability measure on $\mathbb{R}^d$. We call a hyperplane H with $P(H) > 0$ a linear component of P. P is a distribution without linear components, if $P(H) = 0$ for all hyperplanes H.*

$x : \mathbb{R}^d \to \mathbb{R}^d$ denotes the identity on $\mathbb{R}^d$, $x_1, \ldots, x_d$ its components. P is a fixed Borel distribution on $\mathbb{R}^d$ without linear components. We assume that

$$\int \|x\| dP < \infty. \tag{12}$$

DEFINITION 4. *Let us call a partition $\mathcal{B} = (B_1, \ldots, B_m)$ of $\mathbb{R}^d$ a polytopepartition (PT-partition), if for all $i \neq j \in \{1, \ldots, m\}$ a halfspace E exists, such that $B_i \subseteq E$ und $B_j \subseteq \overline{E^c}$. $\tilde{\mathcal{B}}_m$ denotes the set of PT-partitions $\mathcal{B}$ of size at most m.*

$\mathcal{B} = (B_1, \ldots, B_m)$ is called MSP-partition, if linear functions $g_1, \ldots, g_m$ exist such that

$$B_i \subseteq \{x \mid g_i(x) \geq g_j(x) \text{ for all } j\}. \tag{13}$$

In this case $\mathcal{B}$ is generated by g, where $g(x) = \max\{g_i(x) \mid i \leq m\}$.

Note that PT-partitions consist of sets with boundaries that are subsets of hyperplanes and therefore nullsets. We do not distinguish between partitions which are identical up to boundaries. Moreover, we identify functions which are equal P-a.e.

In the one-dimensional case all PT-partitions are MSP-partitions. They consist of intervals. If $\mathcal{B} = (B_1, \dots, B_m)$ with $B_i =]a_{i-1}, a_i]$ for $i < m$, $B_m =]a_{m-1}, \infty[$ and $a_0 < \cdots < a_{m-1}$, then $\mathcal{B}$ is generated by a function $g \in L_m$. Indeed, let $b_1 < \cdots < b_m$ and let g be continuous on $\mathbb{R}$ and linear on B_i with $g' = b_i$ on $\overset{\circ}{B_i}$. g is convex and generates $\mathcal{B}$. For $d > 1$ the class of MSP-partitions is a proper subclass of $\tilde{\mathcal{B}}_m$, cf. Examples 1 and 2.

Let $\mathcal{B} \in \tilde{\mathcal{B}}_m$. Let us denote the quantization defined by (1) and (2) by $\mu^{\mathcal{B}}$. If $g \in L_m$ and $\mathcal{B}$ generated by g, then we abbreviate $\mu^{\mathcal{B}}$ by μ^g. $\mathcal{K}$ denotes the class of continuous, P-integrable convex functions on $\mathbb{R}^d$. $g \in \mathcal{K} \setminus L_{m-1}$ is called nontrivial. We define for $f \in \mathcal{K}$, $\mathcal{O}_f = \{\mu \in \mathcal{M}(P,m) \mid \mu(f) = I_m^f\}$. $\mu \in \mathcal{O}_f$ is called f-optimal.

REMARK 2. (a) In general $|\mathcal{O}_f| > 1$. If $f \in L_m \setminus L_{m-1}$, then $|\mathcal{O}_f| = 1$ and thus $\mathcal{O}_f = \{\mu^f\}$.
(b) Let us mention an implication of the equivalence theorem of Pötzelberger and Strasser [7]: If $f \in \mathcal{K} \setminus L_{m-1}$ and $\mu \in \mathcal{O}_f$, then a $g \in L_m$ exists, such that $g \leq f$, $\mu = \mu^g$ and $\mu(f) = \mu(g) = P(g)$. We call the MSP-partition generated by such a g an f-optimal partition.

Theorem 1 - 4 summarize the essential results on maximal elements of $\mathcal{M}(P,m)$: Existence of admissible quantizations, admissibility of f-optimal quantizations, characterization of admissible quantizations and geometric properties of corresponding partitions. The proofs of these results are provided in sections 4 and 5.

THEOREM 1. *For every admissible $\mu \in \mathcal{M}(P,m)$ there is a $\mathcal{B} \in \tilde{\mathcal{B}}_m$ such that $\mu = \mu^{\mathcal{B}}$.*

THEOREM 2. *For every $\mu \in \mathcal{M}(P,m)$ there is an admissible $\nu \in \mathcal{M}(P,m)$ such that $\mu \preceq \nu$.*

COROLLARY 1. *For every $\mu \in \mathcal{M}(P,m)$ there is a $\mathcal{B} \in \tilde{\mathcal{B}}_m$ with $\mu \preceq \mu^{\mathcal{B}}$.*

THEOREM 3. *Suppose g is a nontrivial integrable convex function. Then all $\mu \in \mathcal{O}_g$ are admissible.*

THEOREM 4. *Let $\mu \in \mathcal{M}(P,m)$ with $|supp(\mu)| = m$. μ is admissible if and only if there exists a sequence of nontrivial convex functions $(g_n)_{n=1}^\infty \subseteq L_m \setminus L_{m-1}$ such that $\mu = \lim_{n\to\infty} \mu^{g_n}$.*

Theorem 1 and Theorem 3 point out the close connection between admissible quantizations and partitions with special geometric properties. Maximizing an information measure $I^g(\mathcal{B})$ with nontrivial g leads to an admissible quantization which is generated by a MSP-partition. On the other hand, Theorem 1 implies that all admissible $\mu \in \mathcal{M}(P,m)$ come from PT-partitions. Thus for

$d = 1$ Theorem 1 and Theorem 3 provide a complete characterization of admissible quantizations as any PT-partition is a MSP-partition. However, for $d > 1$ the situation is different. We have

$$\{\mu^g \mid \mu \in L_m \setminus L_{m-1}\} \subsetneqq \{\mu \mid \mu \text{ admissible in } \mathcal{M}(P,m)\} \subsetneqq \{\mu^{\mathcal{B}} \mid \mathcal{B} \in \tilde{\mathcal{B}}_m\}.$$

Since $\{\mu^g \mid \mu \in L_m \setminus L_{m-1}\}$ is dense in $\{\mu \mid \mu \text{ admissible in } \mathcal{M}(P,m)\}$, in principle any admissible μ may be approximated by a suitable f-optimal quantization. Note that even in the light of the Theorem of Blackwell-Sherman-Stein this proposition is not trivial. If P has linear components Theorem 3 remains valid. However, Theorem 1 and Theorem 4 do not hold, see Pötzelberger [8]. In particular, unbiased quantizations μ and nontrivial convex functions (g_n) may exist, such that μ is not admissible, although $\mu = \lim_{n\to\infty} \mu_n$ with $\mu_n \in \mathcal{O}_{g_n}$.

Let us emphasize an important consequence of Theorem 4 and Theorem 1. All methods that lead to a partition of the sample space which is not the limit of MSP-partitions are inadmissible. This observation applies in particular to various generalizations of principal points or k-means clustering where the expectation of a nonlinear function of the distance to the nearest principal point is minimized. All methods that lead to a disintegration $P = \sum_{i=1}^m w_i P_i$ with overlapping supports of the distributions P_i are inadmissible. All these procedures are dominated by procedures based on admissible quantizations.

3 Examples

We discuss two examples to emphasize that Theorem 4 is not trivial. The first example provides a PT-partition $\mathcal{B}$, which is not a MSP-partition, but the limit of MSP-partions. In the second we construct a PT-partition, which is not even the limit of MSP-partitions.

The Hausdorff metric defines a topology on the set of partitions. More precisely, define for sets $A, C \subset \mathbb{R}^d$,

$$d_{\mathcal{H}}(A,C) = \inf\{\epsilon > 0 \mid A \subseteq C^\epsilon \text{ and } C \subseteq A^\epsilon\}, \tag{14}$$

where the ϵ-neighborhood of a set A is $A^\epsilon = \{x \mid \; y \in A \text{ exists with } \|x-y\| < \epsilon\}$.

DEFINITION 5. *Let $(\mathcal{B}_n) \subseteq \tilde{\mathcal{B}}_m$ be a sequence of PT-partitions, $\mathcal{B}_n = (B_1^n, \ldots, B_m^n)$ and let $\mathcal{B} = (B_1, \ldots, B_{m'}) \in \tilde{\mathcal{B}}_m$. $(\mathcal{B}_n)$ converges to $\mathcal{B}$, ($\mathcal{B} = \lim_{n\to\infty} \mathcal{B}_n$) $:\Leftrightarrow$ For all n a permutation π of $\{1, \ldots, m\}$ exists, such that for all $N > 0$ and $i \leq m'$,*

$$\lim_{n\to\infty} d_{\mathcal{H}}(B_{\pi(i)}^n \cap [-N,N]^d, B_i \cap [-N,N]^d) = 0.$$

EXAMPLE 1. Let $d = 2$, $\mathcal{B} = (B_1, B_2, B_3)$ with $B_1 = [0,\infty[\times\mathbb{R}$, $B_2 =]-\infty,0[\times[0,\infty[$ and $B_3 =]-\infty,0[\times]-\infty,0[$. Then $\mathcal{B} \in \tilde{\mathcal{B}}_3$. Suppose $\mathcal{B}$ is a MSP-partition generated by $g \in L_3 \setminus L_2$. Then $g(x) = \max\{g^1(x), g^2(x), g^3(x)\}$ with g^i linear and $g^i(x) \geq g^j(x)$ for $j \leq 3$ on B_i. W.l.g.

$g^1 = 0$. Since $\bar{B}_1 \cap \bar{B}_2 = \{0\} \times [0, \infty[$ and $\bar{B}_1 \cap \bar{B}_3 = \{0\} \times]-\infty, 0]$, we have for $x = (x_1, x_2)$, $g^2(x) = b_2 x_1$ and $g^3(x) = b_3 x_1$ with $b_i < 0$ and $b_2 \neq b_3$. But then

$$\bar{B}_2 \cap \bar{B}_3 =]-\infty, 0] \times \{0\} \not\subseteq \{g^2 = g^3\} = \{0\} \times \mathbb{R}.$$

To show that $\mathcal{B}$ is the limit of MSP-partitions, choose $g_n = \max\{g_n^1(x), g_n^2(x), g_n^3(x)\}$ with

$$\begin{aligned} g_n^1(x) &= 0, \\ g_n^2(x) &= -x_1 + x_2/n, \\ g_n^3(x) &= -x_1 - x_2/n. \end{aligned}$$

The MSP-partitions generated by (g_n) converge to $\mathcal{B}$. Note that for any distribution P without linear component and with $P(B_i) > 0$ for $i = 1, 2, 3$, $\mu^{\mathcal{B}}$ is admissible in $\mathcal{M}(P, 3)$. The admissibility of $\mu^{\mathcal{B}}$ for a class of distributions can thus be deduced from geometrical properties of $\mathcal{B}$.

EXAMPLE 2. Let $d = 2$ and $\mathcal{B} = (B_1, B_2, B_3, B_4)$. The boundaries of the sets B_i are drawn as thick lines in Figure 1. $U = (0,0)$, $V = (1,0)$, $W = (0,1)$, $R = (-1,-1)$, $S = (2, -1/2)$ and $Q = (q, 2)$ are points on these boundaries. It will be shown that for all $q \neq -1/2$, the partition $\mathcal{B}$ is not the limit of MSP-partitions. We assume that P is a distribution without linear components and that all sets B_i have positive probability.

Suppose, $g_n \in L_4 \setminus L_3$ exist with $\mathcal{B}^n \to \mathcal{B}$, where $\mathcal{B}^n = (B_1^n, B_2^n, B_3^n, B_4^n)$ is the MSP-partition generated by g_n. We have $g_n(x) = \max\{g_n^1(x), g_n^2(x), g_n^3(x), g_n^4(x)\}$, g_n^i linear and $g_n^i > g_n^j$ for $j \neq i$ on $\mathring{B}_i^n$.

In Figure 1 a partition $\mathcal{B}^n$ is drawn with thin lines. $U_n (= U), V_n (= V), W_n (= W), R_n, S_n, Q_n$ denote points on the boundaries of the sets B_i^n which are either points of intersection of three boundaries or points of intersection of boundaries and $\mathbb{R} \times \{-1\}$, $\mathbb{R} \times \{2\}$ or $\{2\} \times \mathbb{R}$. Note that if a sequence (g_n) with $\mathcal{B}^n \to \mathcal{B}$ exists, it can be chosen such that $g_n^1 = 0$, $U_n = U$, $V_n = V$ and $W_n = W$. Furthermore g_n can be multiplied by a constant to get $g_n^2((1/2, 1/2)) = 1/2$. But then $\{x \mid g_n^2(x) = g_n^1(x)\}$ and $g_n^2((1/2, 1/2)$ are determined and therefore

$$g_n^2(x) = x_1.$$

$\{x \mid g_n^2(x) = g_n^3(x)\}$ is determined and $R_n \to R$ implies

$$g_n^3(x) = x_1 + b_n x_2$$

with $b_n \to -1$. Analoguously, the knowlegde of $\{x \mid g_n^2(x) = g_n^4(x)\}$ and $S_n \to S$ imply,

$$g_n^4(x) = c_n(1 - x_2) + (1 - c_n) x_1$$

with $c_n \to -1$. From $\{Q_n\} = \{g_n^4 = g_n^1\} \cap (\mathbb{R} \times \{2\})$ we conclude

$$Q_n = (c_n/(1 - c_n), 2).$$

$c_n \to -1$ and $Q_n \to Q$ imply

$$Q = (-1/2, 2).$$

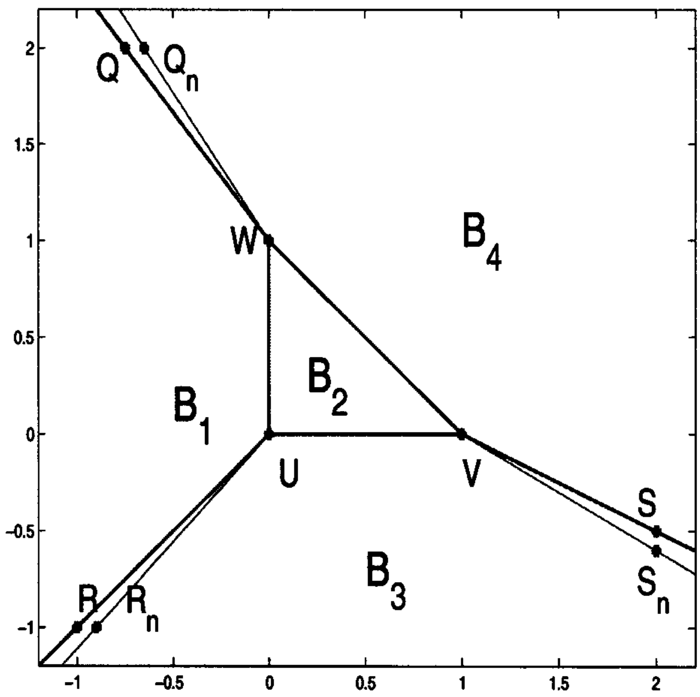

Figure 1: Example 2

4 Auxilliary Results and Proofs

LEMMA 1. *Let P be a probability measure with $\int x dP = 0$. Then for every $w \in]0,1[$ there exists a halfspace H, such that $P(H) = w$ and for $i = 2, \ldots, d$,*

$$\int_H x_i dP = \int_{H^c} x_i dP = 0. \tag{15}$$

PROOF. Let $v \in \mathbb{R}^d$, $\beta \in \mathbb{R}$, $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$. We define

$$H_0(v, \beta) = \{x \mid \langle x - \beta e_1, v \rangle = 0\},$$

$$H^+(v, \beta) = \{x \mid \langle x - \beta e_1, v \rangle > 0\},$$

$$H^-(v, \beta) = \{x \mid \langle x - \beta e_1, v \rangle < 0\}.$$

The proof of the lemma is split into three steps.

Step 1. We prove that for fixed β a $v \in \mathbb{R}^d$ exists with

$$\int_{H^+(v,\beta)} x_i dP = 0 \tag{16}$$

for $i \geq 2$. Note that $\int_{H^-(v,\beta)} x_i dP = 0$ follows then from $\int x dP = 0$. If $P(H^+(v, \beta)) = 1$ we set $\mathbb{E}(x \mid H^-(v, \beta)) = 0$. In this case the statement holds. Therefore assume that for all $v \in \mathbb{R}^d$, $P(H^+(v, \beta)) \in]0,1[$.

The proof is based on the Theorem of Borsuk (see Lorentz et al. [6]). Let $S^d = \{v \mid \|v\| = 1\}$ denote the sphere. The theorem states that for all continuous $\varphi : S^d \to \mathbb{R}^{d-1}$ at least one $v \in S^d$ with $\varphi(v) = \varphi(-v)$ exists.

Define for $v \in S^d$

$$\varphi(v) = (\mathbb{E}(x_2 \mid H^+(v,\beta)), \ldots, \mathbb{E}(x_d \mid H^+(v,\beta))).$$

φ is continuous as hyperplanes have probability 0. Let $v \in S^d$ witht $\varphi(v) = \varphi(-v)$. $H^+(-v,\beta) = H^-(v,\beta)$ implies for $i \geq 2$, $\mathbb{E}(x_i \mid H^+(v,\beta)) = \mathbb{E}(x_i \mid H^-(v,\beta))$. Thus $\int x dP = 0$ yields for $i \geq 2$

$$\begin{aligned} 0 &= \mathbb{E}(x_i) \\ &= \mathbb{E}(x_i \mid H^+(v,\beta))P(H^+(v,\beta)) + \mathbb{E}(x_i \mid H^-(v,\beta))P(H^-(v,\beta)) \\ &= \mathbb{E}(x_i \mid H^+(v,\beta)). \end{aligned}$$

Step 2. Here we prove that for fixed β the sets $H^+(v,\beta))$ which satisfy (16) are unique up to sets of measure 0. More precisely, let $v^1, v^2 \in \mathbb{R}^d$, $\beta \in \mathbb{R}$ with $v_1^1 > 0$, $v_1^2 > 0$ and $\mathbb{E}(x_i \mid H^+(v^1,\beta)) = \mathbb{E}(x_i \mid H^+(v^2,\beta)) = 0$ for all $i \geq 2$. Then

$$P(H^+(v^1,\beta) \setminus H^+(v^2,\beta)) = P(H^+(v^2,\beta) \setminus H^+(v^1,\beta)) = 0. \tag{17}$$

Moreover, $P(H^+(v^1,\beta)) = P(H^+(v^2,\beta))$.

It is sufficient to prove the assertion for $v_1^1 = v_1^2 = 1$ and $\beta = 0$. For notational convenience we abbreviate $H^+(v^1,\beta)$ by H_1^+ and $H^+(v^2,\beta)$ by H_2^+. $\int_{H_1^+} x_i dP = \int_{H_2^+} x_i dP = 0$ implies for $i \geq 2$

$$\int_{H_1^+ \setminus H_2^+} x_i dP = \int_{H_2^+ \setminus H_1^+} x_i dP. \tag{18}$$

Considering the definition of H_1^+ we obtain $\int_{H_1^+ \setminus H_2^+} \langle x, v^1 \rangle dP \geq 0$ and $\int_{H_2^+ \setminus H_1^+} \langle x, v^1 \rangle dP \leq 0$. Therefore

$$\int_{H_1^+ \setminus H_2^+} x_1 dP - \int_{H_2^+ \setminus H_1^+} x_1 dP = \int_{H_1^+ \setminus H_2^+} \langle x, v^1 \rangle dP - \int_{H_2^+ \setminus H_1^+} \langle x, v^1 \rangle dP \geq 0. \tag{19}$$

Analogously the inequality

$$\int_{H_1^+ \setminus H_2^+} x_1 dP - \int_{H_2^+ \setminus H_1^+} x_1 dP = \int_{H_1^+ \setminus H_2^+} \langle x, v^2 \rangle dP - \int_{H_2^+ \setminus H_1^+} \langle x, v^2 \rangle dP \leq 0 \tag{20}$$

holds. From (19) and (20) we see that

$$\int_{H_1^+ \setminus H_2^+} x_1 dP - \int_{H_2^+ \setminus H_1^+} x_1 dP = 0, \tag{21}$$

$$\int_{H_1^+ \setminus H_2^+} \langle x, v^1 \rangle dP = \int_{H_2^+ \setminus H_1^+} \langle x, v^1 \rangle dP = \int_{H_1^+ \setminus H_2^+} \langle x, v^2 \rangle dP = \int_{H_2^+ \setminus H_1^+} \langle x, v^2 \rangle dP = 0. \tag{22}$$

(22) implies $P(H_1^+ \setminus H_2^+) = P(H_2^+ \setminus H_1^+) = 0$.

Step 3. If a halfspace $H^+(v, \beta)$ satisfies (16) then $v_1 \neq 0$. Choose v such that $v_1 > 0$ and define $w(\beta) = P(H^+(v, \beta))$. $\beta \mapsto w(\beta)$ is continuous. Lemma 1 has been proved once we have shown that $\lim_{\beta\to\infty} w(\beta) = 0$: For reasons of symmetry $\lim_{\beta\to-\infty} w(\beta) = 1$ holds and then for all $w \in]0, 1[$ a halfspace H exists with $P(H) = w$.

Therefore, let (β_n) and (v^n) be sequences with $\beta_n \to \infty$, $\|v^n\| = 1$, $v_1^n > 0$ and $H_n^+ := H^+(v^n, \beta_n)$. We may assume that $v \in \mathbb{R}^d$ and $t \in [0, \infty]$ exist with $v^n \to v$ and $v_1^n \beta_n \to t$.

$x \in H_n^+$ implies $\beta_n v_1^n < \langle x, v^n \rangle \leq \|x\|$. It follows that $(w(\beta_n))$ is a null sequence if $t = \infty$.

Assume $t \in [0, \infty[$. Define $y = \sum_{i=2}^d v_i x_i$ and $y_n = \sum_{i=2}^d v_i^n x_i$. From $\mathbb{E}(I_{\{y_n > (\beta_n - x_1)v_1^n\}} y_n) = 0$ we have

$$\mathbb{E}(I_{\{y>t\}} y) = \mathbb{E}((I_{\{y>t\}} - I_{\{y_n > (\beta_n - x_1)v_1^n\}})y) + \mathbb{E}(I_{\{y_n > (\beta_n - x_1)v_1^n\}})(y - y_n))$$

and

$$\mathbb{E}(I_{\{y>t\}} y) \leq \liminf_{n\to\infty} \Big(\mathbb{E}(|I_{\{y>t\}} - I_{\{y_n > (\beta_n - x_1)v_1^n\}}| y) + \mathbb{E}(|y - y_n|)\Big) = 0.$$

In particular, $P(y > t) = 0$. Define $H^+ = \{x \mid y > t\}$. For $\epsilon > 0$ we choose $r > 0$ and $s > 0$, such that for $K_r = \{x \mid \|x\| \leq r\}$ and $D_s = \{x \mid t - s < y \leq t\}$, $P(K_r^c) < \epsilon/2$ and $P(D_s) < \epsilon/2$ hold. If $x \in K_r \cap D_s^c \cap (H_n^+ \setminus H^+)$, then $y \leq t - s$, $\|x\| \leq r$, $y_n > (\beta_n - x_1)v_1^n$ and thus

$$\begin{aligned} y_n - y &> (\beta_n - x_1)v_1^n - (t - s) \\ &\geq s + (\beta_n v_1^n - t) - r v_1^n. \end{aligned}$$

$v_1^n \to 0$, $\beta_n v_1^n \to t$ and $y - y_n \leq \|x\| \|v^n - v\|$ give

$$P(H_n^+) = P(H_n^+ \setminus H^+) \leq P(K_r^c) + P(D_s) + P(\|x\| \|v^n - v\| \geq s + (\beta_n v_1^n - t) - r v_1^n).$$

This shows $\lim_{n\to\infty} P(H_n^+) = 0$ and completes the proof of Lemma 1. □

PROOF of Theorem 1. W.l.g. we may assume $\int x dP = 0$. We split this proof into three parts. In step 1 and 2 we prove that for any $\mu \in \mathcal{M}(P, 2)$ a $\mathcal{B} \in \tilde{\mathcal{B}}_2$ exists such that $\mu \preceq \mu^{\mathcal{B}}$.

Step 1. Let $d = 1$ and $m = 2$. Let $\mu \in \mathcal{M}(P, 2)$, $\mu = w_1 \delta_{p_1} + w_2 \delta_{p_2}$, $P = w_1 P_1 + w_2 P_2$, with $w_1 + w_2 = 1$, $0 < w_i < 1$, $p_1 < 0 < p_2$ and $\int x dP_i = p_i$. Define $s \in \mathbb{R}$ by $P(] - \infty, s]) = w_1$. Furthermore, let $\tilde{P}_1, \tilde{P}_2 << P$ and $\tilde{p}_1, \tilde{p}_2, \tilde{\mu}$ with

$$\begin{aligned} \frac{d\tilde{P}_1}{dP} &= \frac{1}{w_1} I_{]-\infty, s]}, \\ \frac{d\tilde{P}_2}{dP} &= \frac{1}{w_2} I_{]s, \infty[}, \\ \tilde{p}_i &= \int x d\tilde{P}_i, \\ \tilde{\mu} &= w_1 \delta_{\tilde{p}_1} + w_2 \delta_{\tilde{p}_2}. \end{aligned}$$

From $P = w_1\tilde{P}_1 + w_2\tilde{P}_2$ we obtain $\tilde{\mu} \in \mathcal{M}(P,2)$.

To prove $\mu \preceq \tilde{\mu}$, note that $\tilde{P}_1$ is stochastically smaller than P_1 and $\tilde{P}_2$ stochastically larger than P_2. Therefore,

$$\tilde{p}_1 \leq p_1 < p_2 \leq \tilde{p}_2. \tag{23}$$

It is easy to see that (23) implies $\mu \preceq \tilde{\mu}$.

Step 2. The proof of the analogous result for general dimensions relies on Lemma 1. Let $d > 1$ and $m = 2$. Let $\mu \in \mathcal{M}(P,2)$, $\mu = w_1\delta_{p_1} + w_2\delta_{p_2}$. We may assume that p_1 and p_2 are scalar multiples of the unit vector $e_1 = (1,0,\ldots,0)$, i.e. $p_i = a_i e_1$ with $a_i \in \mathbb{R}$. Let the first component of p_1 be negative and that of p_2 positive. Furthermore, we write $P = w_1P_1 + w_2P_2$ as a mixture of distributions P_1 and P_2 with means p_1 and p_2.

According to Lemma 1 a halfspace H exists such that $P(H) = w_1$ and the means $\tilde{p}_1 := \mathbb{E}(x \mid H)$ and $\tilde{p}_2 := \mathbb{E}(x \mid H^c)$ are likewise scalar multiples of e_1, i.e. $\tilde{p}_i = \tilde{a}_i e_1$. Moreover, we may choose H such that $\tilde{a}_1 < 0$ and $\tilde{a}_2 > 0$. Define

$$\tilde{\mu} = w_1\delta_{\tilde{y}_1} + w_2\delta_{\tilde{y}_2},$$

$\tilde{P}_1, \tilde{P}_2 << P$ with

$$\frac{d\tilde{P}_1}{dP} = \frac{1}{w_1} I_H$$

and

$$\frac{d\tilde{P}_2}{dP} = \frac{1}{w_2} I_{H^c}.$$

Let b denote a vector orthogonal to ∂H, the boundary of H, with nonnegative first component. The law of $\langle b, x\rangle$ under $\tilde{P}_1$ is stochastically smaller than under P_1 and under $\tilde{P}_2$ it is stochastically larger than under P_2. Therefore the the distance $\|\tilde{p}_1 - \partial H\|$ is larger than the distance of p_1 from ∂H and similarly $\|\tilde{p}_2 - \partial H\| \geq \|p_2 - \partial H\|$. Consequently

$$\tilde{a}_1 \leq a_1 \leq 0 \leq a_2 \leq \tilde{a}_2.$$

The remainder of the proof of $\mu \preceq \tilde{\mu}$ proceeds along the lines of step 1.

Step 3. We prove the theorem by induction on m. Let $m > 2$. Assume that for all distributions P' without linear component and $\int \|x\| dP' < \infty$ the following statement holds: Let $m' < m$ and let $\mu = \sum_{i=1}^{m'} w_i\delta_{p_i} \in \mathcal{M}(P, m')$ be admissible with $p_i = \int x dP'_i$ and $P' = \sum_{i=1}^{m'} w_i P'_i$. Then for all $i < j \leq m'$ a halfspace H exists, such that $P'_i(H) = P'_j(H^c) = 1$.

Let $\mu \in \mathcal{M}(P,m)$ be admissible with $P = \sum_{i=1}^{m} w_i P_i$ according to (11). Suppose a $i < j \leq m$ exists, such that no halfspace H with $P_i(H) = P_j(H^c) = 1$ exists. Let $k \notin \{i,j\}$,

$$\begin{aligned} P &= w_kP_k + (1-w_k)\sum_{n\neq k} \frac{w_n}{1-w_k} P_n \\ &=: w_kP_k + (1-w_k)P'. \end{aligned}$$

$$\mu' := \sum_{n\neq k} \frac{w_n}{1-w_k}\delta_{p_i}$$

is thus not admissible in $\mathcal{M}(m-1,P')$. Let $\nu' \in \mathcal{M}(m-1,P')$, $\nu' \neq \mu'$ with $\mu' \preceq \nu'$. It is easy to see that $\nu := w_k\delta_{p_k} + (1-w_k)\nu'$ is in $\mathcal{M}(P,m)$, $\mu \preceq \nu$ and $\mu \neq \nu$. □

LEMMA 2. *Let $\mu = \sum_{i=1}^m w_i\delta_{p_i} \in \mathcal{M}(P,m)$. Then for all $i \leq m$*

$$\|p_i\| \leq \int \|x\| dP/w_i. \tag{24}$$

LEMMA 3. *Suppose $E \subseteq \mathbb{R}^d$ is convex and closed with $P(E) = 1$ and $P(\partial E) = 0$. Let (A_n) be a sequence of Borel sets and $p_n = \mathbb{E}(x \mid A_n)$. No accumulation point of (p_n) is in $[-\infty,\infty]^d \setminus \overset{\circ}{E}$ if $\liminf_{n\to\infty} P(A_n) > 0$.*

PROOF. Let $E = \mathbb{R}^d$. The claim follows from $|\mathbb{E}(x \mid A_n)| \leq \int \|x\| dP/P(A_n)$. For the complete proof of the lemma see Pötzelberger und Strasser [7], Lemma 5.2. □

LEMMA 4. *For any increasing sequence (μ_n) in $\mathcal{M}(P,m)$ there exists a smallest upper bound in $\mathcal{M}(P,m)$.*

PROOF. Let $(\mu_n)_{n=1}^\infty$ be an increasing sequence in $\mathcal{M}(P,m)$, i.e. $\mu_n \preceq \mu_{n+1}$ for $n \in \mathbb{N}$. Let $\mu_n = \sum_{i=1}^m w_i^n \delta_{p_i^n}$. Lemma 2 implies that weights of unbounded sequences of prototypes decrease to zero. Thus upon rearranging indices, there exist a suitable subsequence $(n_k) \subseteq \mathbb{N}$, $m' \leq m$, w_i and p_i with $w_i^{n_k} \to w_i$, $p_i^{n_k} \to p_i$ for $i \leq m'$ and $w_i^{n_k} \to 0$ for $i > m'$. Let $\mu_\infty = \sum_{i=1}^{m'} w_i\delta_{p_i}$. Note that $\mu_n \to^{w^*} \mu$ implies $\mu_n(f) \to \mu(f)$ for convex and piecewiese linear functions f: Let $f = \ell + g$ with ℓ linear and g nonnegative. Let $N > 0$ and $g_N = (g-N)_+$. $g - g_N$ is bounded, g_N is convex and therefore $\mu_n(g_N) \leq P(g_N)$, which can be made arbitrarily small for large N.

Thus for convex and piecewise linear functions f and $k', n \in \mathbb{N}$ with $n_{k'} \geq n$ we have

$$\int f d\mu_\infty = \lim_{k\to\infty} \int f d\mu_{n_k} \geq \int f d\mu_{n_{k'}} \geq \int f d\mu_n.$$

Hence μ_∞ is an upper bound of the sequence (μ_n). Standard arguments reveal that μ_∞ is indeed the smallest upper bound. □

PROOF of Theorem 2. Let $\mathcal{F}$ denote a countable set of nonnegative convex and piecewise linear functions such that for all μ', $\mu'' \in \mathcal{M}(P,m)$ with $\mu' \preceq \mu''$ and $\mu' \neq \mu''$ a $f \in \mathcal{F}$ exists with $\int f d\mu' < \int f d\mu''$. Furthermore assume $P(f) \leq 1$ for all $f \in \mathcal{F}$.

Suppose $\mu \in \mathcal{M}(P,m)$ is not admissible. We define for ordinal numbers $\alpha < \aleph_1$, pairs (μ_α, f_α) with $\mu_\alpha \in \mathcal{M}(P,m)$ and $f \in \mathcal{F}$:

1. Let $\alpha = \beta + 1$ be a successor number. If μ_β is admissible, then let $\mu_\alpha = \mu_\beta$ and $f_\alpha \in \mathcal{F}$ arbitrary. Otherwise pick μ_α and $f_\alpha \in \mathcal{F}$ such that

$$\int f_\alpha d\mu_\alpha - \int f_\alpha d\mu_\beta > \frac{1}{2}\sup\{\int f d\mu' - \int f d\mu_\beta \mid \mu' \in \mathcal{M}(P,m), \mu_\beta \preceq \mu', f \in \mathcal{F}\} \tag{25}$$

2. Let α be a limit number. μ_α is then the smallest upper bound of $\{\mu_\beta \mid \beta < \alpha\}$ and f_α is arbitrary.

We will show that an α_0 exists with $\mu_\alpha = \mu_{\alpha_0}$ for all $\alpha \geq \alpha_0$. μ_{α_0} is then admissible.

By contradiction, assume that $\mu_\alpha \preceq \mu_{\alpha+1}$ and $\mu_\alpha \neq \mu_{\alpha+1}$ for all $\alpha < \aleph_1$. Define for $f \in \mathcal{F}$ and $k \in \mathbb{Z}$,

$$\Omega_f^k = \{\alpha \mid f_{\alpha+1} = f \text{ and } 2^k < \int f d\mu_{\alpha+1} - \int f d\mu_\alpha \leq 2^{k+1}\}.$$

$f \in \mathcal{F}$ and $k \in \mathbb{Z}$ exist with Ω_f^k uncountable. In particular Ω_f^k contains $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$. As a result of the definition of $(\mu_{\alpha_1+1}, f_{\alpha_1+1})$ for all $f \in \mathcal{F}$

$$\int f d\mu_{\alpha_4+1} - \int f d\mu_{\alpha_1} < 2(\int f d\mu_{\alpha_1+1} - \int f d\mu_{\alpha_1}) \leq 2^{k+2}.$$

Thus

$$\begin{aligned}
2^{k+2} &\geq \int f d\mu_{\alpha_4+1} - \int f d\mu_{\alpha_1} \\
&= \int f d\mu_{\alpha_4+1} - \int f d\mu_{\alpha_4} + \int f d\mu_{\alpha_4} - \int f d\mu_{\alpha_3+1} \\
&+ \int f d\mu_{\alpha_3+1} - \int f d\mu_{\alpha_3} + \cdots \\
&\vdots \\
&+ \int f d\mu_{\alpha_1+1} - \int f d\mu_{\alpha_1} \\
&\geq \sum_{i=1}^{4} (\int f d\mu_{\alpha_i+1} - \int f d\mu_{\alpha_i}) \\
&> 4 \times 2^k = 2^{k+2},
\end{aligned}$$

a contradiction. □

PROOF of Theorem 3. Let $\mu \in \mathcal{O}_g$ with g nontrivial. In particular, $|\text{supp}(\mu)| = m$. Let us define $\tilde{g} \leq g$ by

$$\tilde{g}(x) = \max\{g(p_i) + \langle x - p_i, d_i \rangle \mid 1 \leq i \leq m\}$$

with $d_i \in D(g, p_i)$. Note that $\mu(\tilde{g}) = \mu(g) = I_m^g$. $\tilde{g} \leq g$ implies $I_m^{\tilde{g}} \leq I_m^g$ and thus $\mu \in \mathcal{O}_{\tilde{g}}$ and $I_m^{\tilde{g}} = I_m^g$. We have

$$\mu^{\tilde{g}}(\tilde{g}) = I_m^{\tilde{g}} = \mu(\tilde{g}) = \mu(g) = I_m^g.$$

From $\tilde{g} \leq g$ we conclude that $\mu^{\tilde{g}} \in \mathcal{O}_g$. Therefore $|\text{supp}(\mu^{\tilde{g}})| = m$ and $\tilde{g} \in L_m \setminus L_{m-1}$. However, if $\tilde{g} \in L_m \setminus L_{m-1}$, then $|\mathcal{O}_{\tilde{g}}| = 1$ and thus $\mu = \mu^{\tilde{g}}$. To prove that μ is admissible, let $\nu \in \mathcal{M}(P, m)$ with $\mu \preceq \nu$. Since $\mu(\tilde{g}) = I_m^{\tilde{g}}$ we have $\nu(\tilde{g}) = I_m^{\tilde{g}}$ and thus $\nu \in \mathcal{O}_{\tilde{g}} = \{\mu\}$. □

5 Proof of Theorem 4

5.1 Step 1: Distributions with bounded support

Here we establish the most involved part of the proof. Let us assume that the support of P is compact. Let $E \subseteq \mathbb{R}^d$ denote a compact and convex set with $P(E) = 1$. We will show that for any admissible $\mu \in \mathcal{M}(P, m)$ a sequence of nontrivial convex functions (g_n) exist such that $\mu = \lim_{n\to\infty} \mu^{g_n}$. We define

$$\mathcal{K}_0 = \{f : E \to [0,1] \mid f \text{ convex on } E\}. \tag{26}$$

We endow $\mathcal{K}_0$ with the topology of pointwise convergence on $\overset{\circ}{E}$.

LEMMA 5. *Let $(f_n), f \in \mathcal{K}_0$. Then the following assertions are equivalent:*

1. $\lim_{n\to\infty} f_n(x) = f(x)$ *for all* $x \in \overset{\circ}{E}$,

2. $\lim_{n\to\infty} \int \|f_n - f\| dP = 0$,

3. $f_n \to f$ *uniformly on all compact* $F \subseteq \overset{\circ}{E}$.

PROOF. The only nontrivial part of the lemma is 2. $\Rightarrow$ 3. Let $f_n, f \in \mathcal{K}_0$ with $\int \|f_n - f\| dP \to 0$ and let $F \subseteq \overset{\circ}{E}$ be compact. Since F is contained in a finite union of compact and convex subsets of $\overset{\circ}{E}$ it is sufficient to verify the claim for convex and compact F. Note that

$$\inf_{x\in F, y\notin E} \|x - y\| > 0.$$

Therefore, a $M > 0$ existst, such that for all $n \in \mathbb{N}$, $x \in F$ and $d_n(x) \in D(f_n, x)$, $\|d_n(x)\| \leq M$. Note that since the functions f_n are convex, $f_n(x) - f_n(y) \leq \langle y - x, d_n(x)\rangle$ and $f_n(x) - f_n(y) \geq \langle x - y, d_n(y)\rangle$. Therefore (f_n) restricted to F is uniformly bounded and equicontinuous. Thus, by the Theorem of Arzela-Ascoli, (f_n) converges uniformly on F to its limit, which is certainly f. $\square$

LEMMA 6. *If $g_n, g \in \mathcal{K}_0$, $\mathcal{B}_n \in \tilde{B}_m$, $\nu, \nu_n \in \mathcal{M}(P, m)$ with $\nu_n = \mu^{\mathcal{B}_n}$, $g_n \to g$ and $\nu_n \to \nu$, then*

$$\nu_n(g_n) \to \nu(g), \tag{27}$$

and

$$I_m^{g_n} \to I_m^g. \tag{28}$$

PROOF. Let $\nu = \sum_{i=1}^{m'} w_i \delta_{p_i}$ and $\nu_n = \sum_{i=1}^{m} w_i^n \delta_{p_i^n}$. For any subsequence (n_k) a further subsequence $(n'_k) \subseteq (n_k)$ exists, such that $(p_i^{n'_k})$ and $(w_i^{n'_k})$ converge. Denote its limits by p'_i and w'_i respectively. If $p'_i \notin \overset{\circ}{E}$, then $w'_i = 0$, see Lemma 3. Let $F \subseteq \overset{\circ}{E}$ compact with $\{p'_i \mid i \leq m\} \cap \overset{\circ}{E} \subseteq F$.

Therefore $\{p'_i \mid i \leq m\} \cap \overset{\circ}{E} = \{p_i \mid i \leq m'\}$. (g_n) converges on F uniformly to g. Furthermore, $\nu_{n'_k}(F) \to 1$ and $\nu_{n'_k}(gI_F) \to \nu(g)$. Thus

$$\begin{aligned} & \lim_{k\to\infty} |\nu_{n'_k}(g_{n'_k}) - \nu(g)| \\ \leq \; & \lim_{k\to\infty} |\nu_{n'_k}(gI_F) - \nu(g)| + \lim_{k\to\infty} \nu_{n'_k}(|g_{n'_k} - g|I_F) + \lim_{k\to\infty} \nu_{n'_k}(I_{F^c}) = 0. \end{aligned}$$

(28) is an immediate consequence of (27). □

LEMMA 7. *Let $g_n \in (L_m \setminus L_{m-1}) \cap \mathcal{K}_0$, $g \in \mathcal{K}_0$, $\nu \in \mathcal{M}(P,m)$ with $g_n \to g$ and $\mu^{g_n} \to \nu$. Then $\mu^g \preceq \nu$. In particular $\nu \in \mathcal{O}_g$.*

PROOF. Note that $g \in L_m$. Let $\mathcal{B}_n$ and $\mathcal{B} = (B_1, \dots, B_{m'})$ denote the MSP-partitions of g_n and g. W.l.g. assume that $\mathcal{B}_n$ converges to a partition $\mathcal{B}^* = (B_1^*, \dots, B_{m''}^*) \in \tilde{\mathcal{B}}_m$. Then we have $\nu = \mu^{\mathcal{B}^*}$. The fact that g is linear on the elements of $\mathcal{B}$ implies that $\mathcal{B}^*$ is finer than $\mathcal{B}$, i.e. for all $i \leq m'$ a suitable $A \subseteq \{1, \dots, m''\}$ exists with $\bar{B}_i = \cup_{j\in A} \bar{B}_j^*$. Consequently $\mu^g \preceq \nu$. □

LEMMA 8. *Let $\mu \in \mathcal{M}(P,m)$ be admissible and U an open neighborhood of μ. Then a constant $c > 0$ exists, such that for all $\nu \in \mathcal{M}(P,m) \setminus U$ there is a g in $\mathcal{K}_0$ with*

$$\mu(g) \geq \nu(g) + c.$$

PROOF. Suppose, on the contrary, that $\nu_n \in \mathcal{M}(P,m) \cap U^c$ exist with $c_n \to 0$, where

$$c_n := \sup\{\mu(g) - \nu_n(g) \mid g \in \mathcal{K}_0\}.$$

W.l.g. let $\nu_n \to \nu$. Thus ν in U^c and a $g \in \mathcal{K}_0$ exists with $c_0 := \mu(g) - \nu(g) > 0$. Since $c_n \geq \mu(g) - \nu_n(g)$ we have

$$0 \geq \lim_{n\to\infty} (\mu(g) - \nu_n(g)) = \mu(g) - \nu(g) = c_0 > 0,$$

a contradiction. □

For the remaining part of step 1 let $\mu = \sum_{i=1}^m w_i \delta_{p_i}$ denote a fixed admissible quantization in $\mathcal{M}(P,m)$. Then $|\mathrm{supp}(\mu)| = m$. We assume that no sequence (g_n) in $(L_m \setminus L_{m-1}) \cap \mathcal{K}_0$ exists with $\mu^{g_n} \to \mu$. This assumption will lead to a contradiction.

There is a neighborhood U_1 of μ with $U_1 \cap \{\mu^g \mid g \in (L_m \setminus L_{m-1}) \cap \mathcal{K}_0\} = \emptyset$. Since $|\mathrm{supp}(\mu)| = m$ a neighborhood U_2 of μ exists such that $U_2 \cap \{\mu^g \mid g \in L_{m-1} \cap \mathcal{K}_0\} = \emptyset$. Thus for $U = U_1 \cap U_2$ we have

$$U \cap \{\mu^g \mid g \in L_m \cap \mathcal{K}_0\} = \emptyset.$$

According to Lemma 8 a constant $c^* > 0$ exists, such that for all $g \in L_m \cap \mathcal{K}_0$ an $h \in \mathcal{K}_0$ exists with

$$\mu(h) - \mu^g(h) \geq c^*. \tag{29}$$

Define for $g \in L_m \cap \mathcal{K}_0$,

$$H(g) = \{h \in \mathcal{K}_0 \mid \mu(h) - \mu^g(h) \geq c^*\}. \tag{30}$$

LEMMA 9. *(a) $H(g)$ is convex and compact.*
(b) If $g, g_n \in L_m \cap \mathcal{K}_0$, $h_n \in H(g_n)$ with $g_n \to g$ and $h_n \to h$, then $h \in H(g)$.
(c) If $h \in H(g)$, $\tilde{h} \in \mathcal{K}_0$ with $\tilde{h} \leq h$ and $\mu(\tilde{h}) = \mu(h)$, then $\tilde{h} \in H(g)$.

PROOF. It suffices to prove (b). Let $g_n, g \in L_m \cap \mathcal{K}_0$, $h_n, h \in \mathcal{K}_0$ with $h_n \in H(g_n)$ and $h_n \to h$. We may assume that (μ^{g_n}) converges to a limit ν. According to Lemma 7, $\nu \in \mathcal{O}_g$ and $\mu^g \preceq \nu$. Since $\mu^g(h) \leq \nu(h)$ and $\mu^{g_n}(h_n) \to \nu(h)$ (see Lemma 6) we have

$$\begin{aligned} \mu(h) - \mu^g(h) &\geq \mu(h) - \nu(h) \\ &= \lim_{n\to\infty} \mu(h_n) - \mu^{g_n}(h_n) \geq c^*. \square \end{aligned}$$

REMARK 3. Note that if a $g \in L_m \cap \mathcal{K}_0$ exists with $g \in H(g)$, then $\mu(g) \geq \mu^g(g) + c^* \geq \mu(g) + c^*$ leads to a contradiction. It is tempting to try to apply the fixpoint theorem of Kakutani to the mapping $g \mapsto H(g)$ with g restricted to a suitable domain. The difficulty of this approach is the choice of this domain. $L_m \cap \mathcal{K}_0$ is not convex. If the choice is $\mathcal{K}_0$ with corresponding definition of the sets $H(g)$, then it has to be verified that $g \mapsto H(g)$ is a so-called K-mapping, i.e. that Lemma 9(*b*) holds. However, $\mathcal{O}_g$ may contain more than a single element. For a proof of a version of Lemma 9(b) it would be necessary to establish the existence of a continuous selection $S : \mathcal{K}_0 \to \mathcal{M}(P, m)$ with $S(g) \in \mathcal{O}_g$. We conjecture that such a selection does not exist in general. Therefore, we apply the fixpoint theorem of Kakutani to a parametrization of suitable support functionals of $L_m \cap \mathcal{K}_0$.

Define for $h \in \mathcal{K}_0$ and $d_i \in D(h, p_i)$ a function $\tilde{h}$ by $\tilde{h}(x) = \max\{h(p_i) + \langle x - p_i, d_i\rangle \mid i \leq m\}$. Then $\tilde{h} \in H(g)$ if $h \in H(g)$. Let $\mathcal{G}_0 \subseteq L_m$ denote the set $\mathcal{G}_0 = \{\tilde{g} \mid g \in \mathcal{K}_0\}$. Consider the following mapping $T : \mathbb{R}^{m(d+1)} \to \mathcal{K}_0$, $z \mapsto T_z$,

$$T_z(x) = \max\{c_i + \langle x - p_i, d_i\rangle \mid i \leq m\} \vee 0, \tag{31}$$

where $z = (c_1, \ldots, c_m, d_1, \ldots, d_m) \in \mathbb{R}^{m(d+1)}$ with $c_1, \ldots, c_m \in \mathbb{R}$, $d_1, \ldots, d_m \in \mathbb{R}^d$. Furthermore, let

$$\Xi = \{z \in \mathbb{R}^{m(d+1)} \mid T_z \in \mathcal{G}_0 \text{ und } c_i = T_z(p_i)\}. \tag{32}$$

LEMMA 10. *With c_i, d_i and $z \in \Xi$ defined above we have*
(a) $0 \leq c_i \leq 1$.
(b) $c_i \geq \max\{c_j + \langle p_i - p_j, d_j\rangle \mid j \neq i\}$.
(c) Ξ is convex.
(d) Ξ is compact.

PROOF. The only nontrivial claim is that Ξ is bounded. Let us prove this claim. Let $z = (c_1, \dots, c_m, d_1, \dots, d_m) \in \mathbb{R}^{m(d+1)}$ with $\tilde{g} = T_z \in \mathcal{G}_0$. According to (a) the components c_i are in $[0,1]$. The subdifferentials d_i are bounded, since the prototypes p_i are in $\overset{\circ}{E}$ and T_z is bounded.

To be more specific, let $i_0 \leq m$ with $\|d_i\| \leq \|d_{i_0}\|$ for all $i \leq m$. Since $p_i \in \overset{\circ}{E}$, we have $\delta > 0$, where

$$\delta := \min\{\|p_i - x\| \mid x \in \partial E, i \leq m\}.$$

Let $t_0 > 0$ with $p_{i_0} + t_0 d_{i_0} \in \partial E$. Then $t_0\|d_{i_0}\| \geq \delta$. Since $\tilde{g}$ is convex, we have

$$\tilde{g}(p_{i_0} + t_0 d_{i_0}) \geq \tilde{g}(p_{i_0}) + t_0\|d_{i_0}\|^2.$$

$1 \geq \tilde{g}(p_{i_0} + t_0 d_{i_0})$, $\tilde{g}(p_{i_0}) \geq 0$ and $t_0\|d_{i_0}\| \geq \delta$ imply $\|d_{i_0}\| \leq 1/\delta$. $\square$

Define for $g \in L_m \cap \mathcal{K}_0$,

$$\tilde{H}(g) = \{z \in \Xi \mid T_z \in H(g)\}. \tag{33}$$

LEMMA 11. *(a) $\tilde{H}(g)$ is compact and nonvoid.*
(b) If $g, g_n \in L_m \cap \mathcal{K}_0$, $z_n \in \Xi$ with $g_n \to g$ and $z_n \to z$, then $z \in \tilde{H}(g)$.
(c) $\tilde{H}(g)$ is convex.

PROOF. (a) and (b) follow from Lemma 9. $\tilde{H}(g)$ is nonvoid, since $h \in H(g)$ implies $\tilde{h} \in H(g)$.

To establish (c), let $z = (c_1, \dots, d_m)$ and $z' = (c_1', \dots, d_m')$ in $\tilde{H}(g)$. For $0 \leq \alpha \leq 1$ we have

$$\begin{aligned} T_{\alpha z + (1-\alpha)z'}(x) &= \max\{\alpha c_i + (1-\alpha)c_i' + \langle x - p_i, \alpha d_i + (1-\alpha)d_i'\rangle \mid i \leq m\} \vee 0 \\ &\leq \alpha \max\{c_i + \langle x - p_i, d_i\rangle \mid i \leq m\} \vee 0 + (1-\alpha)\max\{c_i' + \langle x - p_i, d_i'\rangle \mid i \leq m\} \vee 0 \\ &= \alpha T_z + (1-\alpha)T_{z'}. \end{aligned}$$

Since $H(g)$ is convex, we have $\alpha T_z + (1-\alpha)T_{z'} \in H(g)$.

$$T_{\alpha z + (1-\alpha)z'}(p_i) = (\alpha T_z + (1-\alpha)T_{z'})(p_i)$$

and Lemma 9(c) yield $T_{\alpha z + (1-\alpha)z'} \in H(g)$. $\square$

The set-valued mapping

$$\begin{aligned} \xi : \quad & \Xi \to \Xi \\ z \quad & \mapsto \tilde{H}(T_z) \end{aligned}$$

satisfies the assumptions of the fixpoint theorem of Kakutani (cf. Smart [9]). Therefore a $z \in \Xi$ exists such that $z \in \tilde{H}(T_z)$. Let $g = T_z$. Then we have $g \in L_m \cap \mathcal{K}_0$ and $g \in H(g)$, i.e.

$$\begin{aligned} \mu(g) &\geq \mu^g(g) + c^* \\ &\geq \mu(g) + c^*, \end{aligned}$$

a contradiction.

5.2 Step 2: Distributions with unbounded support

Here we prove that an admissible $\mu \in \mathcal{M}(P, m)$ is the limit $\mu = \lim_{n\to\infty} \mu^{g_n}$, with (g_n) nontrivial, without assuming a bounded support of P.

Let $\mu \in \mathcal{M}(P, m)$ be admissible, N large enough, such that $P(E_N) > 0$, where $E_N := [-N, N]^d$. Denote by P_N the conditionalal distribution given E_N, i.e. $P_N(A) = P(A \cap E_N)/P(E_N)$. According to Theorem 1 a partition $\mathcal{B} = (B_1, \dots, B_m)$ exists with $\mu = \mu^{\mathcal{B}}$. Let

$$\mu^N = \sum_{i=1}^{m} w_i^N \delta_{p_i^N} \in \mathcal{M}(P_N, m)$$

be generated by $\mathcal{B}$ and P_N, i.e. $w_i^N = P_N(B_i)$ and $p_i^N = \mathbb{E}^{P_N}(x \mid B_i)$. Let $\nu_N \in \mathcal{M}(P_N, m)$ be admissible with $\mu^N \preceq \nu^N$. According to Step 1 there are nontrivial convex functions g_n^N such that

$$\nu_N = \lim_{n\to\infty} \tilde{\mu}^{g_n^N},$$

where $\tilde{\mu}^{g_n^N}$ is computed with P_N. Corresponding quantizations computed from P are denoted by $\mu^{g_n^N}$. W.l.g. assume that a ν exists with $\lim_{N\to\infty} \nu_N = \nu$. Then $\nu \in \mathcal{M}(P, m)$ and $\mu \preceq \nu$. Consequently, $\nu = \mu$. Thus for a suitable sequence (n_N),

$$\mu = \lim_{N\to\infty} \tilde{\mu}^{g_{n_N}^N}.$$

The fact that $P_N(B)$ and $\mathbb{E}^{P_N}(x \mid B)$ converge to $P(B)$ and $\mathbb{E}(x \mid B)$ uniformly on the set of polytopes, which are intersections of $m-1$ halfspaces and have a probabiltity bounded from below by a $\delta > 0$, implies that

$$\lim_{N\to\infty} (\tilde{\mu}^{g_{n_N}^N} - \mu^{g_{n_N}^N}) = 0.$$

Thus $\mu = \lim_{N\to\infty} \mu^{g_N}$ with $g_N = g_{n_N}^N$.

5.3 Step 3

In the remainder of this section we established the admissibility of limits $\mu = \lim_{n\to\infty} \mu^{g_n}$ if $g_n \in L_m \setminus L_{m-1}$ and $|\mathrm{supp}(\mu)| = m$. We do so by induction on m.

Let $m = 2$. Any limit $\mu = \lim_{n\to\infty} \mu^{g_n}$ is itself of the form $\mu = \mu^g$ with a nontrivial g and is thus admissible.

Let therefore $m > 2$ and assume that for all distributions $\tilde{P}$ without linear component and with $\int \|x\| d\tilde{P} < \infty$, and all $m' < m$, $\mu' \in \mathcal{M}(\tilde{P}, m')$ is admissible if $|\mathrm{supp}(\mu')| = m'$ and if $\mu' = \lim_{n\to\infty} \mu^{g_n'}$, with $(g_n') \subseteq L_{m'} \setminus L_{m'-1}$.

Assume $(g_n) \subseteq L_m \setminus L_{m-1}$ and $\mu = \lim_{n\to\infty} \mu^{g_n}$. Let $\nu \in \mathcal{M}(P, m)$ be admissible with $\mu \preceq \nu$. Partitions $\mathcal{B}^n = (B_1^n, \dots, B_m^n)$ and $\mathcal{B}^* = (B_1^*, \dots, B_m^*)$ exist such that $\nu = \mu^{\mathcal{B}^*}$ and $\mu^{g_n} = \mu^{\mathcal{B}^n}$. We may assume that $(\mathcal{B}^n)$ converges to a limit $\mathcal{B} = (B_1, \ldots, B_n)$. Then $\mu = \mu^{\mathcal{B}}$. Let $\mu = \sum_{i=1}^m w_i \delta_{p_i}$, $\mu^{g_n} = \sum_{i=1}^m w_i^n \delta_{p_i^n}$ and $\nu = \sum_{i=1}^m w_i^* \delta_{p_i^*}$. Furthermore, let $g_n(x) = \max\{g_i^n(x) \mid i \leq m\}$ with

$$g_i^n(x) = g_i^n(p_i^n) + \langle x - p_i^n, d_i^n \rangle,$$

and $d_i^n \in \mathbb{R}^d$. Denote $E_N = [-N, N]^d$. We choose $N > 0$ large enough, such that for a suitable $c_0 > 0$ and all i, n:

1. $p_i, p_i^n, p_i^* \in \overset{\circ}{E}_N$,

$$d(p_i, \partial(E_N \cap \bar{co}(\mathrm{supp}(P) \cap B_i))) \geq c_0,$$

$$d(p_i^n, \partial(E_N \cap \bar{co}(\mathrm{supp}(P) \cap B_i^n))) \geq c_0,$$

$$d(p_i^*, \partial(E_N \cap \bar{co}(\mathrm{supp}(P) \cap B_i^*))) \geq c_0.$$

2. $P(B_i \cap E_N) \geq P(B_i)/2$, $P(B_i^n \cap E_N) \geq P(B_i^n)/2$ and $P(B_i^* \cap E_N) \geq P(B_i^*)/2$.

($\bar{co}(A)$ denotes the closed convex hull of A and $d(x, A) := \inf\{\|x - y\| \mid y \in A\}$).

W.l.g. the functions g_n may be chosen such that

(a) $g_m^n = 0$,

(b) $0 \leq g_n \leq 1$ on $E_N \cap \mathrm{supp}(P)$,

(c) $\max\{g_n(x) \mid x \in E_N \cap \mathrm{supp}(P)\} = \max\{g_n^1(x) \mid x \in E_N \cap \mathrm{supp}(P)\} = 1$,

(d) a $g \in L_m \cap \mathcal{K}$ exists such that $g_n \to g$.

Note that g may be trivial. Given assumption 1. on E_N, the set $\{\|d_i^n\| \mid i \leq m, n \in \mathbb{N}\}$ is bounded. Let $\hat{\mathcal{B}} = (\hat{B}_1, \ldots, \hat{B}_{m'})$ with $m' \leq m$ be the MSP-partition generated by g. Note that g is linear an the sets $\hat{B}_i$. Since for all n, $d(p_1^n, \partial(E_N \cap \bar{co}(\mathrm{supp}(P) \cap B_1^n))) \geq c_0 > 0$ holds, g is not identically zero on $\hat{B}_1$. Thus $g > 0$ on $\overset{\circ}{\hat{B}}_1$. Furthermore, as $g = 0$ holds on $\hat{B}_{m'}$, $\hat{\mathcal{B}}$ consists of at least two sets, i.e. $m' \geq 2$.

Assume that $\hat{\mathcal{B}}$ is labelled such that $B_1^n \to A \subseteq \hat{B}_1$ and $B_m^n \to C \subseteq \hat{B}_{m'}$.

The fact that $\mu \in \mathcal{O}_g$ and $\mu \preceq \nu$ implies $\nu \in \mathcal{O}_g$. g is therefore linear on all sets B_i and B_i^*. In particular $\mathcal{B}$ and $\mathcal{B}^*$ are finer than $\hat{\mathcal{B}}$ (up to boundaries, which are sets of probability zero).

Denote by $\mu_{|\hat{B}_1}$ and $\nu_{|\hat{B}_1}$ the conditional distributions of μ and ν given $\hat{B}_1$ and by $\mu_{|\hat{B}_1^c}$, $\nu_{|\hat{B}_1^c}$ those given $\hat{B}_1^c$. Furthermore, let $\nu = \sum_{i=1}^m w_i K(. \mid p_i)$ with $\int x K(dx \mid p_i) = p_i$. Since $\mu, \nu \in \mathcal{O}_g$, $p_i \in \hat{B}_1$ implies $p_i \in \overset{\circ}{\hat{B}}_1$ and $\mathrm{supp}(K(. \mid p_i)) \subseteq \bar{\hat{B}}_1$. Moreover, $p_i^* \in B_i^*$ implies $p_i^* \in \overset{\circ}{B_i^*}$ such that even

$$\mathrm{supp}(K(. \mid p_i)) \subseteq \overset{\circ}{\hat{B}}_1$$

if $p_i \in \hat{B}_1$. Similarily,

$$\mathrm{supp}(K(. \mid p_i)) \subseteq \overset{\circ}{\hat{B}_1^c}$$

if $p_i \notin \hat{B}_1$. This reveals that

$$\mu_{|\hat{B}_1} \preceq \nu_{|\hat{B}_1}, \tag{34}$$

and

$$\mu_{|\hat{B}_1^c} \preceq \nu_{|\hat{B}_1^c}. \tag{35}$$

Let $m_1 = |\mathrm{supp}(\mu_{|\hat{B}_1})|$ and $k_1 = |\mathrm{supp}(\nu_{|\hat{B}_1})|$.

Suppose, $m_1 \geq k_1$. Note that

$$\mu_{|\hat{B}_1} = \lim_{n\to\infty} \mu^{\tilde{g}_n} \in \mathcal{M}(P_{|\hat{B}_1}, m_1), \tag{36}$$

where $\tilde{g}_n(x) = \max\{g_i^n(x) \mid p_i^n \in \hat{B}_1\}$.

Since $P_{|\hat{B}_1}$ is a distribution without linear component, (34) and (36) imply $\mu_{|\hat{B}_1} = \mu_{|\hat{B}_1}$. Therefore, $k_1 = m_1$. Similar arguments applied to $\mu_{|\hat{B}_1^c}$ and $\nu_{|\hat{B}_1^c}$ lead to $\mu_{|\hat{B}_1^c} = \nu_{|\hat{B}_1^c}$. Therefore $\mu = \nu$ holds.

If $m_1 \leq k_1$, then $m - m_1 = |\mathrm{supp}(\mu_{|\hat{B}_1^c})| \leq m - k_1 = |\mathrm{supp}(\nu_{|\hat{B}_1^c})|$. In this case we may prove first $\mu_{|\hat{B}_1^c} = \nu_{|\hat{B}_1^c}$ and finally $\mu_{|\hat{B}_1} = \nu_{|\hat{B}_1}$.

REFERENCES

[1] D. Blackwell. Comparison of experiments. In L. LeCam and J. Neyman, eds., *Proc. 2nd Berkeley Symp. Math. Statistics Prob.*, 93-102, 1951.

[2] D. Blackwell. Equivalent comparisons of experiments. *Ann. Math. Statistics 24*, 265-272, 1953.

[3] H.H. Bock. A clustering technique for maximizing ϕ-divergence, noncentrality and discriminating power. In M. Schrader, editor, *Analyzing and Modeling Data and Knowledge*, 19-36, Berlin Heidelberg New York, 1992. Springer Verlag

[4] B. A. Flury. Principal points. *Biometrika 77,* 33-41, 1990.

[5] T. Kohonen. Self-organization and associative memory. Springer, 1984.

[6] G. G. Lorentz, M. v. Golitschek and Y. Makovoz. Constructive Approximations: Advanced Problems. Springer-Verlag, 1996.

[7] K. Pötzelberger and H. Strasser. Clustering and quantization by MSP-Partitions, 2000. To appear in: *Statistics and Decisions.*

[8] K. Pötzelberger. Admissible unbiased quantizations: Distributions with linear components, 2000. Submitted.

[9] D. R. Smart. Fixed point theorems. Cambridge University Press, 1974.

[10] V. Strassen. The existence of probablity measures with given marginals. *Ann. Math. Statist.* *36*, 423-439, 1965.

[11] H. Strasser. Towards a statistical theory of optimal quantization. Submitted 2000.

[12] E. Torgersen. Comparison of statistical experiments. Cambridge University Press, 1991.